

# Distributed Index Coding

Parastoo Sadeghi<sup>†</sup>, Fatemeh Arbabjolfaei\*, and Young-Han Kim\*

<sup>†</sup>Research School of Engineering, Australian National University, Canberra, ACT, 2601, Australia

\*Department of Electrical and Computer Engineering, University of California, San Diego, CA 92093, USA

Emails: parastoo.sadeghi@anu.edu.au, {farbabjo, yhk}@ucsd.edu

**Abstract**—In this paper, we study the capacity region of the general distributed index coding. In contrast to the traditional centralized index coding where a single server contains all  $n$  messages requested by the receivers, in the distributed index coding there are  $2^n - 1$  servers, each containing a unique non-empty subset  $J$  of the messages and each is connected to all receivers via a noiseless independent broadcast link with an arbitrary capacity  $C_J \geq 0$ . First, we generalize the existing polymatroidal outer bound on the capacity region of the centralized problem to the distributed case. Next, building upon the existing centralized composite coding scheme, we propose three distributed composite coding schemes and derive the corresponding inner bounds on the capacity region. We present a number of interesting numerical examples, which highlight the subtleties and challenges of dealing with the distributed index coding, even for very small problem sizes of  $n = 3$  and  $n = 4$ .

## I. INTRODUCTION

The index coding problem is a canonical problem in network information theory and has been studied over the past two decades using tools from various disciplines such as combinatorics, algebra, and information theory [1], [2]. In the traditional index coding problem, it is assumed that one server has all messages requested by the receivers. However, in many practical circumstances this assumption might not be true and the messages might be distributed over multiple servers. We refer to this more general version as the *distributed index coding problem* compared to the traditional *centralized index coding problem*. The distributed index coding problem was first studied by Ong, Ho, and Lim [3], where lower and upper bounds on the optimal codelength were derived in the special case in which each message is only known to one receiver and each receiver only knows one message a priori and it is shown that the bounds match if no two servers have any messages in common.

In this paper, we study the distributed index coding problem in its general form, which to the best of our knowledge has not been investigated before. First, we generalize the polymatroidal outer bound on the capacity region of the centralized problem [4] to the distributed case. Next, building upon the existing centralized composite coding scheme [4], we propose three distributed composite coding schemes, derive the corresponding inner bounds on the capacity region and show their use via examples. Although the polymatroidal outer bound is tight for all centralized index coding problems with up to  $n = 5$  messages [4], there exist instances of the distributed problem with  $n = 3$  messages for which the polymatroidal outer bound is not tight. Nevertheless, using customized Shannon-type inequalities, we show that the proposed distributed composite coding scheme achieves the

capacity region for all instances of the problem with  $n = 3$  messages.

In this paper,  $[n]$  denotes the set  $\{1, 2, \dots, n\}$  and the set of all nonempty subsets of  $[n]$  is  $N \doteq \{J \subseteq [n]: J \neq \emptyset\}$ .

## II. SYSTEM MODEL AND PROBLEM SETUP

Consider the following index coding problem. There are  $n$  messages in the system,  $x_1, x_2, \dots, x_n$ , where  $x_j \in \{0, 1\}^{t_j}$  for  $j \in [n]$  and some  $t_j$ . There are  $n$  receivers, where receiver  $j$  wants to obtain message  $x_j$  and knows a subset of the messages a priori, denoted by  $x(A_j)$  for some  $A_j \subseteq [n] \setminus \{j\}$ . For simplicity of notation throughout the paper and where the context is clear, we will refer to  $j$  as the wanted message and to  $A_j$  as the side information of receiver  $j$ , respectively. Any instance of this problem can be specified by a side information graph  $G$  with  $n$  nodes, in which a directed edge  $i \rightarrow j$  represents that receiver  $j$  has message  $i$  as side information ( $i \in A_j$ ). For instance, Fig. 1 shows the directed graph representing the index coding problem with  $A_1 = \emptyset$ ,  $A_2 = \{3\}$ , and  $A_3 = \{2\}$ .

The main difference in the system model compared to traditional (centralized) index coding problem is in the server setup. Instead of a single server which contains all messages, there are  $2^n - 1$  servers. For each  $J \in N$ , there is a server that contains all messages  $j \in J$  and the capacity of the broadcast link connecting server  $J$  to all receivers is denoted by  $C_J$ . Hence, we assume that there are  $2^n - 1$  ideal bit pipes to the receivers with arbitrary link capacities. This is a fairly general model that allows for all possible message availabilities on different servers. If  $C_J = 1$  only for  $J = [n]$  and is zero otherwise, we recover the centralized index coding problem. A special normalized symmetric case is where  $C_J = 1$  for all  $J \in N$ . Server  $J$  sends sequence  $y_J \in \{0, 1\}^{u_J}$  for some  $u_J$  to all receivers which is a function of the messages at that server.

Based on the side information  $A_j$  and the received bits  $y_J \in \{0, 1\}^{u_J}$  from all servers, receiver  $j$  finds the estimate  $\hat{x}_j$  of the message  $x_j$  (see Fig. 2). We say that rate-capacity

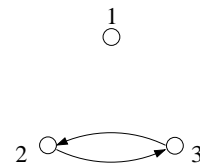


Fig. 1. A 3-node side information graph representing the index coding problem with  $A_1 = \emptyset$ ,  $A_2 = \{3\}$ , and  $A_3 = \{2\}$ .

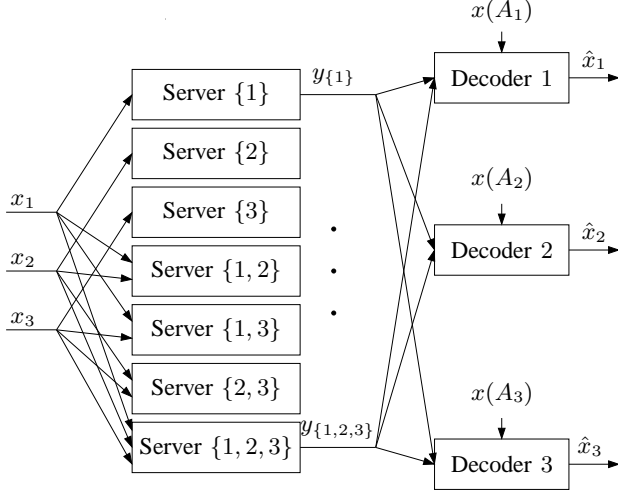


Fig. 2. The distributed index coding problem with  $n = 3$ .

tuple  $(\mathbf{R}, \mathbf{C}) = ((R_j, j \in [n]), (C_J : J \in N))$  is achievable if there exists  $r$  such that:

$$R_j \leq \frac{t_j}{r}, \quad C_J \geq \frac{u_J}{r}, \quad \forall j \in [n], \quad \forall J \in N. \quad (1)$$

For a given  $\mathbf{C}$ , the capacity region  $\mathcal{C}$  of this index coding problem is the closure of the set of achievable rate tuples  $\mathbf{R} = (R_1, \dots, R_n)$ .

Throughout the paper, we will compactly represent the distributed index coding problem (for a given  $\mathbf{C}$ ) as sets of  $(j|i \in A_j)$ . For example, for  $A_1 = \emptyset$ ,  $A_2 = \{3\}$ , and  $A_3 = \{2\}$ , we write  $(1); (2|3); (3|2)$ .

### III. OUTER BOUNDS

In this section, we generalize the polymatroidal outer bound for the centralized index coding problem (see, for example, [4, Theorem 1]) to the distributed case.

*Theorem 1:* Let  $B_j$  be the set of interfering messages at receiver  $j$ , i.e.,  $B_j = [n] \setminus (A_j \cup \{j\})$ . If  $(\mathbf{R}, \mathbf{C})$  is achievable, then for every  $T \in N$ ,

$$R_j \leq f_T(B_j \cup \{j\}) - f_T(B_j), \quad j \in T, \quad (2)$$

for some  $f_T(S)$ ,  $S \subseteq T$ , such that

- 1)  $f_T(\emptyset) = 0$ ,
- 2)  $f_T(T) = \sum_{J: J \cap T \neq \emptyset} C_J$ ,
- 3)  $f_T(A) \leq f_T(B)$  for all  $A \subseteq B \subseteq T$ , and
- 4)  $f_T(A \cup B) + f_T(A \cap B) \leq f_T(A) + f_T(B)$ ,  $\forall A, B \subseteq T$ .

A relaxed version of this bound, which is a generalized version of the maximal acyclic induced subgraph (MAIS) bound, is handy and also useful for some problems.

*Corollary 2:* If  $(\mathbf{R}, \mathbf{C})$  is achievable for an index coding problem represented by the directed graph  $G$ , then for every  $T \in N$  it must satisfy

$$\sum_{j \in S} R_j \leq \sum_{J: J \cap T \neq \emptyset} C_J, \quad (3)$$

for all  $S \subseteq T$  for which the subgraph of  $G$  induced by  $S$  does not contain a directed cycle.

*Example 3:* Consider the index coding problem shown in Fig. 1. The following is the generalized MAIS outer bound for this problem (inactive inequalities are not shown).

$$\begin{aligned} R_1 &\leq C_{\{1\}} + C_{\{1,2\}} + C_{\{1,3\}} + C_{\{1,2,3\}}, \\ R_2 &\leq C_{\{2\}} + C_{\{1,2\}} + C_{\{2,3\}} + C_{\{1,2,3\}}, \\ R_3 &\leq C_{\{3\}} + C_{\{1,3\}} + C_{\{2,3\}} + C_{\{1,2,3\}}, \\ R_1 + R_2 &\leq C_{\{1\}} + C_{\{2\}} + C_{\{1,2\}} + C_{\{1,3\}} + C_{\{2,3\}} + C_{\{1,2,3\}}, \\ R_1 + R_3 &\leq C_{\{1\}} + C_{\{3\}} + C_{\{1,2\}} + C_{\{1,3\}} + C_{\{2,3\}} + C_{\{1,2,3\}}. \end{aligned}$$

### IV. INNER BOUNDS

In this section, we present a series of inner bounds on the capacity region of the distributed index coding problem, all of which are built upon the method of composite index coding, first introduced in [4]. The first inner bound is a simple extension of [4], in which we solve the composite coding problem separately for each server. We show the shortcoming of this method through an example and then present a new composite coding scheme that solves the problem collectively across all servers. The limitation of this scheme, which is assigning the same decoding sets across all servers, is demonstrated through an example. This leads us to the last inner bounding scheme as a generalization of the two previous methods, where we allow a general grouping of servers for solving distributed index coding.

In this section, we assume that for each non-empty subset  $K \subseteq J$  at server  $J \in N$ , there is a virtual encoder at that server with associated composite coding rate  $C_{K,J}$ .

#### A. Combining Centralized Composite Coding Inner Bounds

The idea is to separately apply the composite coding inner bound to each server and appropriately combine the achievable rates. Fix server  $J$  and for  $j \notin J$ , set  $R_j = 0$ . For every  $j \in J$ , let  $D_{j,J}$  be the index of messages that receiver  $j$  wishes to decode from server  $J$  and  $A_{j,J} = A_j \cap J$ . The rates of the composite messages belong to the polymatroidal rate region  $\mathcal{R}(D_{j,J}|A_{j,J})$  defined by

$$\sum_{j \in L_J} R_{j,J} < \sum_{K \subseteq D_{j,J} \cup A_{j,J}: K \cap L_J \neq \emptyset} C_{K,J}, \quad (4)$$

for all  $L_J \subseteq D_{j,J} \setminus A_{j,J}$ . Then the achievable rate region for server  $J$  is given by

$$\mathbf{R}_J \in \bigcap_{j \in J} \bigcup_{D_{j,J} \subseteq J: j \in D_{j,J}} \mathcal{R}(D_{j,J}|A_{j,J}), \quad (5)$$

with the following constraint on composite rates of server  $J$

$$\sum_{K: K \not\subseteq A_j} C_{K,J} \leq C_J, \quad \forall j \in J. \quad (6)$$

After separately finding the composite coding rate regions for all servers, we add the corresponding constraints. Note that we need to write all possible rate equalities and inequalities, including the inactive ones. Table I shows the application of this scheme to the problem  $(1|3); (2|1); (3|2)$  when  $C_J = 1$  for all  $J \in N$ . We note that the sum rate  $R_1 + R_2 + R_3$  is limited to 7.5. However, the outer bound of Theorem 1 gives a sum rate of  $R_1 + R_2 + R_3 < 9$ . In the next subsection,

we will see that this outer bound can be achieved by treating servers collectively.

### B. Distributed Composite Coding with Same Decoding Sets

Here, we fix a decoding set  $D_j$  for receiver  $j$  across all servers and solve a single distributed coding problem. Intuitively, the advantage of this scheme is that receivers can collectively use servers to decode messages, even though some of those messages do not exist on some servers.

Let  $D_j$  be the index of messages receiver  $j$  decodes. The rates of the composite messages belong to the polymatroidal rate region  $\mathcal{R}(D_j|A_j)$  defined by

$$\sum_{j \in L} R_{j,[n]} < \sum_{K \subseteq D_j \cup A_j: K \cap L \neq \emptyset} \sum_{J: K \subseteq J} C_{K,J}, \quad (7)$$

for all  $L \subseteq D_j \setminus A_j$ . The second sum on the RHS is because all servers that contain  $K$  contribute to the composite rate. Similar as before, we have

$$\mathbf{R}_{[n]} \in \bigcap_{j \in [n]} \bigcup_{D_j \subseteq [n]: j \in D_j} \mathcal{R}(D_j|A_j), \quad (8)$$

and the constraints on the composite rates are

$$\sum_{K: K \not\subseteq A_j} C_{K,J} \leq C_J, \quad \forall j \in [n], \forall J \in N. \quad (9)$$

Table II shows the composite coding inner bound along with an optimal decoding set for all non-isomorphic cases with  $n = 3$  and when  $C_J = 1$  for all  $J \in N$ . It can be seen that for the problem  $(1|3); (2|1); (3|2)$  studied in Table I, the inner bound matches the MAIS outer bound, hence establishing the capacity region. In fact, for all non-isomorphic cases shown in Table II, the composite coding inner bound matches the MAIS outer bound, except for the cases that are in the same group as the problem  $(1); (2|3); (3|2)$ . For these instances, the inequality

$$R_1 + R_2 + R_3 \leq 9 \quad (10)$$

is not deducible by Theorem 1. However, using customized Shannon-type inequalities, we show that this inequality is also an outer bound inequality (see Appendix B). Therefore, for all non-isomorphic index coding problems with  $n = 3$ , the rates shown in Table II are indeed capacity regions. Although the capacity regions are only shown for the case of  $C_J = 1$ ,  $J \in N$ , it can be easily verified that the capacity regions of all problems with  $n = 3$  are achieved by the composite coding scheme for arbitrary values of  $C_J$ ,  $J \in N$ . Furthermore, in Table II we can see that the capacity regions of the two problems  $(1); (2|3); (3|2)$  and  $(1|3); (2|3); (3|2)$  are not the same. This proves the important point that, as opposed to the structural property of the centralized index coding [5], [6], [7], removing an edge that does not belong to any directed cycle may change the capacity region of the distributed index coding problem.

To show the limitation of this composite coding scheme, consider the problem  $(1|4); (2|3, 4); (3|1, 2); (4|2, 3)$ . We find

$R_1 + R_2 + R_3 + R_4 < 23$  to be the best achievable sum rate under this scheme with the full inner bound characterized as

$$\begin{aligned} R_1 &< 8, & R_1 + R_2 &< 12 \\ R_2 &< 8, & R_1 + R_3 &< 12 \\ R_3 &< 8, & R_1 + R_4 &< 12 \\ R_4 &< 8, & R_3 + R_4 &< 12 \\ R_1 + R_2 + R_3 &< 18, & R_1 + R_2 + R_3 + R_4 &< 23 \end{aligned}$$

which is determined by setting  $D_1 = \{1\}$ ,  $D_2 = \{1, 2\}$ ,  $D_3 = \{3, 4\}$  and  $D_4 = \{1, 4\}$  as in its centralized index coding counterpart studied in [4]. However, the outer bound of Theorem 1 gives  $R_1 + R_2 + R_3 + R_4 \leq 24$ . In the next subsection, we will show that indeed this sum rate can be achieved and the capacity region can be established by appropriately grouping servers and assigning suitable different decoding sets to each group.

### C. Distributed Composite Coding with Different Decoding Sets

Intuitively, this scheme can be thought of as a generalization of the previous two schemes where index coding was solved by treating servers either separately or collectively. In this new scheme, we jointly optimize grouping of servers and decoding sets across server groupings.

Consider a subset of servers  $P \subseteq N$  and let  $J' = \cup_{J \in P} J$  be the union of all messages held by servers in  $P$ . For  $j \notin J'$ , we set  $R_j = 0$ . For every  $j \in J'$ , let  $D_{j,P}$  be the index of messages that receiver  $j$  wishes to decode from servers in  $P$  and  $A_{j,P} = \cup_{J \in P} (A_j \cap J)$ . The rates of the composite messages belong to the polymatroidal region  $\mathcal{R}(D_{j,P}|A_{j,P})$ :

$$\sum_{j \in L_P} R_{j,P} < \sum_{K \subseteq D_{j,P} \cup A_{j,P}: K \cap L_P \neq \emptyset} \sum_{J: J \in P, K \subseteq J} C_{K,J},$$

for all  $L_P \subseteq D_{j,P} \setminus A_{j,P}$ . Then the achievable rate region for servers in  $P$  is given by

$$\mathbf{R}_P \in \bigcap_{j \in J'} \bigcup_{D_{j,P} \subseteq J': j \in D_{j,P}} \mathcal{R}(D_{j,P}|A_{j,P}), \quad (11)$$

with the following constraint on composite rates

$$\sum_{K: K \not\subseteq A_j} C_{K,J} \leq C_J, \quad \forall j \in J', \forall J \in P. \quad (12)$$

After finding the composite coding rate regions for all server groupings, we appropriately add the corresponding constraints. Note that in this scheme we need to write all possible rate equalities and inequalities, including the inactive ones. Also, note that the previous two composite coding schemes are special cases of this method with  $2^n - 1$  server groups of the form  $P = J$  in Section IV-A and a single server group of  $P = N$  in Section IV-B, respectively.

The scheme is best shown via the example of Table III where we revisit the problem  $(1|4); (2|3, 4); (3|1, 2); (4|2, 3)$ . Each row represents a server grouping and the corresponding decoding sets are shown in the last column. Note for example, how the decoding set for receiver 1 varies across different server groupings. The obtained inner bound matches the MAIS outer bound, hence establishing the capacity region.

TABLE I  
RATE REGION FOR PROBLEM (1|3); (2|1); (3|2) WITH COMBINING CENTRALIZED COMPOSITE CODING.

Server Index $J$	$R_1$	$R_2$	$R_3$	$R_1 + R_2$	$R_1 + R_3$	$R_2 + R_3$	$R_1 + R_2 + R_3$
$\{1\}$	$< 1$	$= 0$	$= 0$	$< 1$	$< 1$	$= 0$	$< 1$
$\{2\}$	$= 0$	$< 1$	$= 0$	$< 1$	$= 0$	$< 1$	$< 1$
$\{3\}$	$= 0$	$= 0$	$< 1$	$= 0$	$< 1$	$< 1$	$< 1$
$\{1,2\}$	$< 1$	$< 1$	$= 0$	$< 1$	$< 1$	$< 1$	$< 1$
$\{1,3\}$	$< 1$	$= 0$	$< 1$	$< 1$	$< 1$	$< 1$	$< 1$
$\{2,3\}$	$= 0$	$< 1$	$< 1$	$< 1$	$< 1$	$< 1$	$< 1$
$\{1,2,3\}$	$< 1$	$< 1$	$< 1$	$< 1$	$< 1$	$< 1$	$< 1.5$
Sum rates	$< 4$	$< 4$	$< 4$	$< 6$	$< 6$	$< 6$	$< 7.5$

TABLE II  
CAPACITY REGION FOR ALL NON-ISOMORPHIC PROBLEMS OF SIZE  $n = 3$ .

IC Problem	$R_1$	$R_2$	$R_3$	$R_1 + R_2$	$R_1 + R_3$	$R_2 + R_3$	$R_1 + R_2 + R_3$	An Optimal Decoding Set
(1); (2); (3) (1 2); (2); (3) (1 2, 3); (2); (3) (1); (2 3); (3 1) (1); (2 1); (3 1) (1); (2 1); (3 1, 2)	$\leq 4$	$\leq 4$	$\leq 4$	$\leq 6$	$\leq 6$	$\leq 6$	$\leq 7$	$D_j = [n] \setminus A_j$
(1 3); (2 1); (3 2)	$\leq 4$	$\leq 4$	$\leq 4$	$\leq 6$	$\leq 6$	$\leq 6$	$\leq 9$	$D_j = [n] \setminus A_j$
(1); (2 3); (3 2) (1); (2 1, 3); (3 2) (1); (2 1, 3); (3 1, 2)	$\leq 4$	$\leq 4$	$\leq 4$	$\leq 6$	$\leq 6$	$\leq 8$	$\leq 9$	If $A_j = \emptyset$ then $D_j = \{j\}$ , otherwise $D_j = [n] \setminus A_j$
(1 3); (2 3); (3 2) (1 3); (2 1, 3); (3 2) (1 2, 3); (2 3); (3 2)	$\leq 4$	$\leq 4$	$\leq 4$	$\leq 6$	$\leq 6$	$\leq 8$	$\leq 10$	$D_j = [n] \setminus A_j$
(1 3); (2 3); (3 1, 2) (1 2, 3); (2 3); (3 1, 2)	$\leq 4$	$\leq 4$	$\leq 4$	$\leq 6$	$\leq 8$	$\leq 8$	$\leq 10$	$D_j = [n] \setminus A_j$
(1 2, 3); (2 1, 3); (3 1, 2)	$\leq 4$	$\leq 4$	$\leq 4$	$\leq 8$	$\leq 8$	$\leq 8$	$\leq 12$	$D_j = [n] \setminus A_j$

TABLE III  
CAPACITY REGION FOR THE PROBLEM (1|4); (2|3, 4); (3|1, 2); (4|2, 3) IS ACHIEVED VIA SERVER GROUPINGS.

Server Groupings	$R_1$	$R_2$	$R_3$	$R_4$	$R_1 + R_2$	$R_1 + R_3$	$R_1 + R_4$	$R_3 + R_4$	Relevant Optimal Decoding Sets
$P = \{\{1\}\}$	$\leq 1$	$= 0$	$= 0$	$= 0$	$\leq 1$	$\leq 1$	$\leq 1$	$= 0$	$D_1 = \{1\}$
$P = \{\{2\}\}$	$= 0$	$\leq 1$	$= 0$	$= 0$	$\leq 1$	$= 0$	$= 0$	$= 0$	$D_2 = \{2\}$
$P = \{\{3\}\}$	$= 0$	$= 0$	$\leq 1$	$= 0$	$= 0$	$\leq 1$	$= 0$	$\leq 1$	$D_3 = \{3\}$
$P = \{\{4\}\}$	$= 0$	$= 0$	$= 0$	$\leq 1$	$= 0$	$= 0$	$\leq 1$	$\leq 1$	$D_4 = \{4\}$
$P = \{\{1, 2\}\}$	$\leq 1$	$\leq 1$	$= 0$	$= 0$	$\leq 1$	$\leq 1$	$\leq 1$	$= 0$	$D_1 = \{1, 2\}, D_2 = \{1, 2\}$
$P = \{\{1, 3\}, \{3, 4\}\}$	$\leq 1$	$= 0$	$\leq 2$	$\leq 1$	$\leq 1$	$\leq 2$	$\leq 2$	$\leq 2$	$D_1 = \{1, 3\}, D_3 = \{3, 4\}, D_4 = \{1, 4\}$
$P = \{\{2, 3\}, \{2, 4\}\}$	$= 0$	$\leq 2$	$\leq 1$	$\leq 1$	$\leq 2$	$\leq 1$	$\leq 1$	$\leq 2$	$D_2 = \{2\}, D_3 = \{3, 4\}, D_4 = \{4\}$
$P = \{\{1, 4\}, \{1, 2, 3, 4\}, \{1, 2, 3\}, \{1, 2, 4\}, \{1, 3, 4\}, \{2, 3, 4\}\}$	$\leq 5$	$\leq 4$	$\leq 4$	$\leq 5$	$\leq 6$	$\leq 6$	$\leq 6$	$\leq 6$	$D_1 = \{1\}, D_2 = \{1, 2\}, D_3 = \{3, 4\}, D_4 = \{1, 4\}$
Sum rates	$\leq 8$	$\leq 8$	$\leq 8$	$\leq 8$	$\leq 12$	$\leq 12$	$\leq 12$	$\leq 12$	

## V. CONCLUSION

We studied a more general class of index coding problems than those considered in the literature, where for a given problem of size  $n$  we allowed all possible distributed servers with arbitrary broadcast link capacities. We derived a generalized polymatroidal outer bound on the capacity region. We also proposed a new general composite coding scheme and showed the necessity of using this general scheme to achieve the distributed index coding capacity region for a problem of size  $n = 4$ . We showed that even for small distributed index coding problems of size  $n = 3$ , there exist instances where the generalized polymatroidal outer bound is not tight and customized Shannon-type inequalities are required to establish the capacity region. Moreover, we showed instances of the distributed index coding problems of size  $n = 3$ , whose graph structural properties are fundamentally different from their centralized index coding counterparts.

There are many interesting directions for future research work on this challenging problem. One direction is to derive tighter outer bounds. The other general direction is to study

the graph structural properties of the problem and other graph-based inner and outer bounds.

## APPENDIX A PROOF OF THEOREM 1

Consider a  $(t_1, \dots, t_n, r)$  index code for the problem  $(j|A_j)$ ,  $j \in [n]$ . Let  $X_j$  be the uniform random variable over  $\{0, 1\}^{t_j}$  corresponding to message  $j \in [n]$  and  $Y_J$  be the uniform random variable over  $\{0, 1\}^{u_J}$  corresponding to the output of server  $J \in N$ . For each  $T \in N$  we set  $x_j = \emptyset$ ,  $j \notin T$ . Therefore,  $Y_J = \emptyset$ , if  $J \cap T = \emptyset$ . Let  $Y \doteq (Y_J, J \cap T \neq \emptyset)$ . For  $j \in T$

$$\begin{aligned} t_j &= H(X_j) = I(X_j; Y | X(A_j \cap T)) \\ &= H(Y | X(A_j \cap T)) - H(Y | X_j, X(A_j \cap T)). \end{aligned}$$

Define

$$f_T(S) = \sum_{J: J \cap T \neq \emptyset} C_J \frac{H(Y | X(S^c \cap T))}{H(Y)}. \quad (13)$$

Then the set function  $f_T$  satisfies  $f_T(\emptyset) = 0$ , and  $f_T(T) = \sum_{J: J \cap T \neq \emptyset} C_J$ . Let  $A \subseteq B \subseteq T$  ( $B^c \subseteq A^c$ ), then  $f_T(A) \leq f_T(B)$  since conditioning reduces entropy. We show that the set function  $f_T$  is submodular. Consider  $A, B \subseteq T$ , then

$$f_T(A \cup B) - f_T(A) = \frac{\sum_{J: J \cap T \neq \emptyset} C_J}{H(Y)} \times (H(Y|X(A^c \cap B^c \cap T)) - H(Y|X(A^c \cap T))), \quad (14)$$

$$f_T(B) - f_T(A \cap B) = \frac{\sum_{J: J \cap T \neq \emptyset} C_J}{H(Y)} \times (H(Y|X(B^c \cap T)) - H(Y|X((A^c \cup B^c) \cap T))). \quad (15)$$

We have

$$\begin{aligned} & H(Y|X(A^c \cap B^c \cap T)) - H(Y|X(A^c \cap T)) \\ &= I(Y; X(A^c \cap B \cap T) | X(A^c \cap B^c \cap T)) \\ &= H(X(A^c \cap B \cap T)) \\ &- H(X(A^c \cap B \cap T) | Y, X(A^c \cap B^c \cap T)), \end{aligned} \quad (16)$$

and similarly

$$\begin{aligned} & H(Y|X(B^c \cap T)) - H(Y|X((A^c \cup B^c) \cap T)) \\ &= H(X(A^c \cap B \cap T)) \\ &- H(X(A^c \cap B \cap T) | Y, X(B^c \cap T)). \end{aligned} \quad (17)$$

Since conditioning reduces entropy, (14)-(17) imply

$$f_T(A \cup B) - f_T(A) \leq f_T(B) - f_T(A \cap B),$$

which completes the proof of the submodularity of the set function  $f_T$  defined in (13). Let  $(\mathbf{R}, \mathbf{C})$  be achievable. Then for every  $T \in N$  and every  $j \in T$  we have

$$\begin{aligned} R_j &\leq \frac{t_j}{r} = (f_T(B_j \cup \{j\}) - f_T(B_j)) \frac{H(Y)}{r \sum_{J: J \cap T \neq \emptyset} C_J} \\ &\leq (f_T(B_j \cup \{j\}) - f_T(B_j)) \frac{\sum_{J: J \cap T \neq \emptyset} u_J}{r \sum_{J: J \cap T \neq \emptyset} C_J} \\ &\leq f_T(B_j \cup \{j\}) - f_T(B_j). \end{aligned}$$

## APPENDIX B

### PROOF OF INEQUALITY (10)

Consider a  $(t_1, t_2, t_3, r)$  index code for the problem (1); (2|3); (3|2). Let  $X_j$  be the uniform random variable over  $\{0, 1\}^{t_j}$  corresponding to message  $j \in \{1, 2, 3\}$  and  $Y_J$  be the uniform random variable over  $\{0, 1\}^{u_J}$  corresponding to the output of server  $J \in N_3 = \{J \subseteq \{1, 2, 3\}, J \neq \emptyset\}$ . Since the output of every server is a function of the messages available at that server,

$$H(Y_J | X_i, i \in J) = 0, \quad J \in N_3, \quad (18)$$

and due to the exact recovery condition at each receiver

$$H(X_1 | Y_J, J \in N_3) = 0, \quad (19)$$

$$H(X_2 | (Y_J, J \in N_3), X_3) = 0, \quad (20)$$

$$H(X_3 | (Y_J, J \in N_3), X_2) = 0. \quad (21)$$

The independence of the messages and (18)-(21) leads to

$$\begin{aligned} H(X_2 | X_1) &= H(X_2 | X_1, X_3) \\ &= H(X_2 | X_1, X_3, Y_{\{1\}}, Y_{\{3\}}, Y_{\{1,3\}}) \\ &= H(X_2, Y_{\{2\}}, Y_{\{1,2\}}, Y_{\{2,3\}}, Y_{\{1,2,3\}} | X_1, X_3, Y_{\{1\}}, Y_{\{3\}}, Y_{\{1,3\}}) \\ &= H(Y_{\{2\}}, Y_{\{1,2\}}, Y_{\{2,3\}}, Y_{\{1,2,3\}} | X_1, X_3, Y_{\{1\}}, Y_{\{3\}}, Y_{\{1,3\}}) \\ &\leq H(Y_{\{2\}}, Y_{\{1,2\}}, Y_{\{2,3\}}, Y_{\{1,2,3\}} | X_1). \end{aligned} \quad (22)$$

Next consider

$$\begin{aligned} I(X_2; Y_{\{2\}}, Y_{\{1,2\}}, Y_{\{2,3\}}, Y_{\{1,2,3\}} | X_1) \\ &= H(X_2 | X_1) - H(X_2 | Y_{\{2\}}, Y_{\{1,2\}}, Y_{\{2,3\}}, Y_{\{1,2,3\}}, X_1) \\ &= H(Y_{\{2\}}, Y_{\{1,2\}}, Y_{\{2,3\}}, Y_{\{1,2,3\}} | X_1) \\ &- H(Y_{\{2,3\}}, Y_{\{1,2,3\}} | X_1, X_2). \end{aligned} \quad (23)$$

Comparing (22) and (23) we get

$$\begin{aligned} H(X_2 | Y_{\{2\}}, Y_{\{1,2\}}, Y_{\{2,3\}}, Y_{\{1,2,3\}}, X_1) \\ \leq H(Y_{\{2,3\}}, Y_{\{1,2,3\}} | X_1, X_2). \end{aligned} \quad (25)$$

We have

$$\begin{aligned} H(X_1, X_2, X_3 | Y_J, J \in N_3) &= H(X_2 | (Y_J, J \in N_3), X_1) \\ &\leq H(Y_{\{2,3\}}, Y_{\{1,2,3\}} | X_1, X_2) \end{aligned} \quad (26)$$

$$\leq H(Y_{\{2,3\}}, Y_{\{1,2,3\}}), \quad (27)$$

where (26) follows from (25).

$$\begin{aligned} t_1 + t_2 + t_3 &= H(X_1, X_2, X_3) \\ &= H(X_1, X_2, X_3) + H(Y_J, J \in N_3 | X_1, X_2, X_3) \\ &= H(Y_J, J \in N_3) + H(X_1, X_2, X_3 | Y_J, J \in N_3) \\ &\leq H(Y_J, J \in N_3) + H(Y_{\{2,3\}}, Y_{\{1,2,3\}}) \\ &\leq \sum_{J \in N_3} u_J + u_{\{2,3\}} + u_{\{1,2,3\}} \\ &\leq r \left( \sum_{J \in N_3} C_J + C_{\{2,3\}} + C_{\{1,2,3\}} \right) = 9r, \end{aligned} \quad (28)$$

where (28) follows from (27) and (29) follows from the assumption of  $C_J = 1, J \in N_3$ . Therefore,  $R_1 + R_2 + R_3 \leq 9$ .

## REFERENCES

- [1] Y. Birk and T. Kol, "Coding on demand by an informed source (ISCOD) for efficient broadcast of different supplemental data to caching clients," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2825–2830, Jun. 2006.
- [2] M. Effros, S. El Rouayheb, and M. Langberg, "An equivalence between network coding and index coding," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 967–971.
- [3] L. Ong, C. K. Ho, and F. Lim, "The single-uniprior index-coding problem: The single-sender case and the multi-sender extension," 2014. [Online]. Available: <http://arxiv.org/abs/1412.1520>
- [4] F. Arbabjolfaei, B. Bandemer, Y.-H. Kim, E. Sasoglu, and L. Wang, "On the capacity region for index coding," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 962–966.
- [5] M. Gadouleau and S. Riis, "Graph-theoretical constructions for graph entropy and network coding based communications," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6703–6717, Oct. 2011.
- [6] M. Tahmasbi, A. Shahrabi, and A. Gohari, "Critical graphs in index coding," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2014, pp. 281–285.
- [7] F. Arbabjolfaei and Y.-H. Kim, "Structural properties of index coding capacity using fractional graph theory," in *Proc. IEEE Int. Symp. Inf. Theory*, Hong Kong, Jun. 2015, pp. 1034–1038.